

# **O comportamento do indivíduo em uma abordagem probabilística em fonologia**

The behavior of the individual in a probabilistic  
approach in phonology

Raquel M. Fontes Martins  
**Universidade Federal de Minas Gerais**

## **Abstract**



This work argues about the variation in the behavior of the individual in a probabilistic approach in phonology, analyzing the applicability of some phonological phenomena between the individuals evaluated here. We argue in accordance to the theoretical referencial that we adopt – Usage Based Phonology (BYBEE, 2001), Exemplars Model (PIERREHUMBERT, 2001) and Probabilistic Linguistics (BOD; HAY; JANNEDY, 2003) – that the organization of the linguistic component is based on multiple representations, linked in nets and regulated by probabilistic criteria. Amongst others, we make an analysis of the words in study and show that words related to a specific phenomenon can have different levels of accomplishment.

## **Keywords**

Individual, Probabilistic approach, Phonology

## **Resumo**

Este trabalho aborda a variação existente no comportamento do indivíduo em uma abordagem probabilística em fonologia, analisando indivíduos a aplicabilidade de fenômenos fonológicos entre os indivíduos aqui avaliados. Argumentamos, em consonância com o



suporte teórico que adotamos – Fonologia de Uso (BYBEE, 2001), Teoria dos Exemplos (PIERREHUMBERT, 2001) e Linguística Probabilística (BOD; HAY; JANNEDY, 2003) –, que a organização do componente lingüístico é baseada em representações múltiplas, alinhadas em redes e reguladas por critérios probabilísticos. Analisamos as palavras em estudo e mostramos que palavras relacionadas a um fenômeno específico podem ter diferentes comportamentos.

### Palavras-chave

Indivíduo; Abordagem probabilística; Fonologia

## 1. Introdução

Neste trabalho, discutimos a variabilidade no comportamento do indivíduo segundo uma abordagem probabilística em fonologia. Para tanto, propomos a análise de três fenômenos fonológicos do português brasileiro. Argumentamos que a organização do componente lingüístico é baseada em múltiplas representações, alinhavadas entre si (em redes) e reguladas por critérios probabilísticos. Dessa forma, adotamos, como suporte teórico, a Fonologia de Uso (BYBEE, 2001), a Teoria de Exemplares (PIERREHUMBERT, 2001) e a Lingüística Probabilística (BOD; HAY; JANNEDY, 2003).

Na próxima seção, abordamos, brevemente, essa base teórica, para, na seção 3, apresentarmos estudos que discutem a respeito do comportamento do indivíduo. A seção 4 aborda a metodologia aqui empregada, enquanto a seção 5 apresenta uma análise feita para as palavras em estudo. Por fim, a seção 6 apresenta as considerações finais deste trabalho.

## 2. Base teórica

Tradicionalmente, considera-se que as representações lingüísticas excluem informações redundantes e operam com categorias discretas. Nessa proposta, as representações lingüísticas são consideradas como simples, e o mapeamento entre as formas abstratas e as formas atestadas nas línguas é complexo (JOHNSON; MULLENIX, 1997). Essa idéia é uma das bases da fonologia moderna e da lingüística em geral.

Essa perspectiva tradicional tem sido questionada e, em contrapartida, algumas teorias – como a Fonologia de Uso (BYBEE, 2001), a Teoria de Exemplares (PIERREHUMBERT, 2001) e a Lingüística Probabilística (BOD; HAY; JANNEDY, 2003) – sugerem que as representações lingüísticas são múltiplas e contêm informações redundantes com categorias gradientes. Neste trabalho, adotamos tais teorias. Dessa forma, assumimos que a organização do componente

fonológico é baseada nessas múltiplas representações de palavras, alinhavadas entre si em redes e reguladas por critérios probabilísticos.

Na figura 1 dos ANEXOS, há uma representação em rede (BYBEE, 2001), a qual apresenta um esquema com conexões lexicais (fonológicas e semânticas – relações morfológicas) para o sufixo nominal agentivo – *dor*,<sup>1</sup> presente em palavras, como *apresentador* e *trabalhador*, na referida figura. Esse esquema demonstra o que dissemos anteriormente sobre a organização do componente fonológico ser baseada em palavras em rede.

A figura 2 (ver ANEXOS) apresenta uma nuvem de exemplares (PIERREHUMBERT, 2001) com múltiplas representações do sufixo –*dor*: [doh], [doR] e [do]. Já a figura 3 (ver ANEXOS) apresenta duas nuvens de exemplares, com as formas [doh] e [do] em competição. Pela figura 3, percebemos que, se em um primeiro momento, a forma plena [doh] apresenta mais exemplares (sendo, portanto, mais freqüente), em um segundo momento, a forma reduzida [do] cresce o seu número de exemplares de tal maneira que passa a tê-los em maior quantidade do que a forma plena [doh].

O conhecimento probabilístico em fonologia é importante a este estudo, visto que, como afirmam Bod; Hay; Jannedy (2003, p. 2, 3, tradução nossa), “indivíduos diferem não somente nas variantes específicas que eles usam em diferentes contextos, mas também na freqüência com que eles as usam. O conhecimento de variação deve envolver o conhecimento de freqüências”. Na próxima seção, discutimos acerca da variabilidade no comportamento do indivíduo, abordando trabalhos que apontam para essa questão.

### 3. Discussão a respeito do comportamento do indivíduo

Trabalhos que abordam mudança e variação sonora, tradicionalmente, não consideram a variabilidade no comportamento do indivíduo. Nas palavras de Hazen (2002, p. 501-502, tradução nossa):

Para alguns pesquisadores, padrões de variação da língua de indivíduos são idioletos. Contudo, de acordo com a análise variacionista tradicional, especificamente Weinreich *et al.* (1968) e Labov (1989b), idioletos não são uma realidade teórica. [...] Nesta visão, a gramática da variação lingüística existe no nível da comunidade de fala, mas não no nível do indivíduo.

Diferentemente dessa visão tradicional abordada por Hazen, pesquisas atuais atestam variações inter e intra-individuais em fenômenos de variação e mudança sonora (Cf. GUIMARÃES, 2004; HUBACK, 2003; LABOV, 2001; OLIVEIRA, 1992; PAIVA; DUARTE, 2003; SANGSTER, 2002), apesar de não assumirem o comportamento do indivíduo como unidade de análise. Marshall (2004, p. 26) aponta para a importância de se estudar a variabilidade no comportamento do indivíduo, quando afirma: “A prática de agrupar informantes para mostrar padrões de variação linguística é, claro, um procedimento válido; contudo, muita variação individual sistemática é desprezada por esse método” (tradução nossa).

Nesse mesmo sentido, Pagotto (2004, p. 49) questiona:

Qual é o papel da dimensão subjetiva no funcionamento da variação, se noções como idioleto e a fala do indivíduo não são, como aponta Labov (1972a) a fonte da variação, mas a comunidade linguística? Ao mesmo tempo, como lidar com fenômenos de variação em que a dimensão subjetiva parece desempenhar um papel crucial [...]?

Questões como essas levantadas por Pagotto norteiam o presente trabalho. Muitos estudos (OLIVEIRA, 1992; SANGSTER, 2002; NARO, 2003; PAIVA; DUARTE, 2003) vêm demonstrando ser importante investigar o comportamento do indivíduo em fenômenos de mudança e variação sonora. Reinecke (a sair, p. 68), abordando a gramática individual, faz a seguinte afirmação: “A localização da gramática e, portanto, da emergência, está no indivíduo e o processo pelo qual emergências nas gramáticas podem estender-se e reproduzir-se através do sistema, é a interação humana”. Essa autora afirma ainda:

[...] a língua [é] como um sistema complexo, composta por formas e organizações dessas formas emergentes e representada na gramática individual. As formas, ou os recursos linguísticos, neste modelo [Fonologia de Uso, de Bybee (2001)], são armazenadas na gramática individual de cada falante. As gramáticas individuais estão permanentemente em processos de adaptação e mudança [...]. A gramática [...] é ancorada no falante individual (REINECKE, a sair, p. 91).

Neste trabalho, compartilhamos dessa noção de gramática individual de Reinecke. Modelos multirrepresentacionais, como os adotados neste estudo (os

quais foram discutidos na seção anterior), permitem a investigação da variabilidade no comportamento do indivíduo. Nesses modelos, o indivíduo é importante, porque ele contribui para a emergência e a dinamicidade dos sistemas, o que se pode apreender pela afirmação de Reinecke, citada anteriormente. Na próxima seção, abordamos a metodologia empregada no presente trabalho.

#### 4. Metodologia

A fim de avaliarmos as variações inter e intra-individuais, propusemos a análise destes três fenômenos fonológicos – casos de mudança sonora no português brasileiro – distribuídos nestes três *corpora*:

- a) **Corpus 1 – cancelamento de “r” em nominais**, por exemplo: calor a calô (HUBACK, 2003);
- b) **Corpus 2 – cancelamento das líquidas “l” e “r” intervocálicas**, por exemplo: óculos, espírito a ócus, espíto (FONTES MARTINS, 2001);
- c) **Corpus 3 – cancelamento de encontro consonantal tautossilábico**, por exemplo: refrigerante a refigerante (CRISTÓFARO-SILVA, 2000a, b, 2003).

Na seleção dos participantes desta pesquisa, consideramos a noção de líder da mudança proposta por Labov (2001, v. 2), visto que o fato de existirem indivíduos líderes e não-líderes em relação a um fenômeno de mudança sonora aponta para a relevância do comportamento do indivíduo. Em linhas gerais, na proposta desse autor, a classificação de um falante como líder ou não-líder se baseia nos índices estatísticos obtidos para aquele falante, em relação a uma variável inovadora que reflete uma mudança sonora. Assim, um líder apresenta maiores índices de realização de uma variável inovadora do que um falante não-líder que apresenta o inverso, ou seja, menores índices para a realização dessa variável.

Considerando essa questão da liderança, selecionamos os dois informantes líderes e os dois informantes não-líderes de cada um dos três *corpora* apresentados, o que totaliza 12 participantes (seis líderes e seis não-líderes).

Esta pesquisa avalia os três fenômenos acima em dois tipos de coleta: a) uma coleta com todos os 12 participantes, para o estudo transversal, b) outras seis coletas com dois (um líder e um não-líder) desses 12 informantes, para o estudo longitudinal. Pesquisas desenvolvidas dentro do Programa de Estudos sobre o Uso da Língua (PEUL) apontam variações inter e intra-individuais em relação

a duas coletas de dados que foram realizadas em épocas distintas (Cf. PAIVA; DUARTE, 2003). Disso, especialmente, decorreu a nossa decisão de fazer o estudo longitudinal.

Três diferentes estilos foram avaliados nesta pesquisa: entrevista de fala espontânea, tarefa de nomeação por figura e tarefa de leitura. Sangster (2002), analisando variações na pronúncia do inglês de Liverpool, encontrou que o estilo de fala foi um fator importante nas variações inter e intra-individuais.

Os dados avaliados neste estudo são os seguintes:

- **dados da coleta do estudo transversal: 30 palavras mais freqüentes** (ver QUADRO 1 dos ANEXOS) → 15 palavras selecionadas dos três *corpora* citados anteriormente (as cinco palavras mais freqüentes de cada um dos *corpora*) + 15 palavras selecionadas do *corpus* LAEL-escrita e fala (as cinco palavras mais freqüentes em relação a cada um dos três fenômenos em análise).
- **dados das seis coletas do estudo longitudinal:** todas as palavras que apresentam contexto fonético para a ocorrência dos três fenômenos aqui focalizados.

Na próxima seção, apresentamos uma breve análise dos dados coletados. Nessa ocasião, faremos uma comparação entre as palavras do estudo transversal e palavras do estudo longitudinal.

## 5. Análise das palavras em estudo

No total, foram coletados 4.955 dados do estudo transversal e 6.598 dados do estudo longitudinal. Tais dados foram submetidos à análise estatística do programa SPSS, v.13.

As figuras 4 e 5 (ver ANEXOS) apresentam uma análise feita por esse programa (Classificação em Árvore – método Chaid), a qual compara as palavras avaliadas no estudo transversal com palavras do estudo longitudinal. A árvore da figura 4 apresenta todos os dados do estudo transversal: 4.955 ocorrências para os 30 tipos de palavras distribuídas em sete nós. Já a árvore da figura 5 mostra alguns dados do estudo longitudinal: mais exatamente, 1.905 ocorrências para 16 tipos de palavras, de um total de 1.646 tipos de palavras com 6.598 ocorrências.

Deve-se esclarecer que, no estudo longitudinal, em razão da grande quantidade de tipos de palavra (1.646 tipos), foi feito um recorte em que se

consideraram somente palavras com cinco ou mais ocorrências, o que totalizou 39 tipos de palavras. Contudo, dessas 39 palavras do estudo longitudinal, apenas 16 coincidiram com palavras do estudo transversal. Assim, na árvore da figura 5, somente essas 16 palavras (agrupadas em cinco nós) foram consideradas.

Duas observações principais podem ser feitas a respeito dessas duas árvores das figuras 4 e 5. Uma primeira observação é que, nas duas árvores, palavras relacionadas a um mesmo fenômeno – como *grande* e *outro* – apresentam índices muito distintos de realização do fenômeno em questão, no caso, o cancelamento de encontro tautossilábico. A palavra *grande* (ver 6º nó da árvore da figura 4) apresenta um baixo índice de 1,1% de cancelamento em média. Ao contrário, a palavra *outro* (ver 7º nó da árvore da figura 4) apresenta um alto índice de 39,8% de cancelamento. Isso indica que os indivíduos analisados no estudo transversal avaliaram de forma diferente palavras de um mesmo fenômeno. Esse fato aponta para a dinamicidade e a plasticidade do léxico, na medida em que palavras relacionadas a um mesmo fenômeno não têm um comportamento homogêneo, como seria esperado, por exemplo, em uma proposta de caráter neogramático (OSTHOFF; BRUGMAN, 1878).

A outra observação a ser feita é que, considerando-se a árvore da figura 5 relativa aos dados do estudo longitudinal, nota-se que o agrupamento da maioria das 16 palavras dessa árvore é parecido com o agrupamento que foi feito para elas na árvore da figura 4 relativa aos dados do estudo transversal. Por exemplo, *agora* e *melhor* (ver 1º nó da figura 4 e 1º nó da figura 5) *eles* e *sempre* (ver 3º nó da figura 4 e 4º nó da figura 5) apresentam índices de cancelamento parecidos nas duas árvores. *Agora* e *melhor* (que se encontram em um mesmo nó nas duas árvores) apresentam um índice médio de cancelamento, enquanto *eles* e *sempre* (que também se encontram em um mesmo nó nas duas árvores) apresentam um índice mais alto de cancelamento. Isso aponta para o fato de que, na organização do componente fonológico, as palavras se relacionam de forma probabilística, já que a questão da frequência (BYBEE, 2001) foi importante para definir a relação entre elas.

## 6. Considerações finais

O estudo do comportamento do indivíduo é de grande relevância para uma compreensão mais ampla da organização do componente lingüístico (OLIVEIRA, 1992; SANGSTER, 2002; NARO, 2003; PAIVA; DUARTE, 2003). Esse componente apresenta um léxico que opera com plasticidade, alinhando



representações múltiplas (em redes) reguladas por critérios probabilísticos (BOD; HAY; JANNEDY, 2003). Assim como as palavras podem ter representações múltiplas, também o indivíduo pode assumir comportamentos lingüísticos diversos. Nesse sentido, o comportamento do indivíduo seria também múltiplo, adaptando-se a usos de fala específicos. Indivíduos se diferem quanto ao léxico e à gramática. Propomos que a variação no comportamento do indivíduo tem importantes implicações teóricas, pois, nessa abordagem probabilística da fonologia, o componente fonológico é considerado como dinâmico, mutável e organizado pela experiência (BYBEE, 2001).

## NOTA

<sup>1</sup> O sufixo nominal agentivo *-dor*, segundo Rocha (1998, p. 113), forma “substantivos-agentivos a partir de verbos: pescador, namorador, franqueador, apelidador [...]”. Como veremos na seção 4, “Metodologia”, palavras com esse sufixo se relacionam a um dos três fenômenos em estudo, o fenômeno do cancelamento de “r” final em nominais (HUBACK, 2003).

## Referências Bibliográficas

- BOD, J.; HAY, J.; JANNEDY, S (Ed.). *Probabilistic Linguistics*. Cambridge: MIT Press, 2003.
- BYBEE, Joan. *Phonology and language use*. Cambridge: Cambridge, 2001.
- CRISTÓFARO-SILVA, Thaís. Sobre a queda dos encontros consonantais no português brasileiro. *Estudos lingüísticos XXIX*. V. 29, 2000a.
- CRISTÓFARO-SILVA, Thaís. Branching Onsets in Brazilian Portuguese. *30<sup>th</sup> Linguistic Symposium in Romance Languages*. University of Florida. Gainesville, 2000b.
- CRISTÓFARO-SILVA, Thaís. Descartando fonemas: a representação mental na Fonologia de Uso In: DA HORA, Dermeval; COLLISCHONN, Gisela (org.). *Teoria Lingüística: Fonologia e outros temas*. Recife: Ed. Universitária.UFPB, 2002, p. 200-231.
- CRISTÓFARO-SILVA, Thaís. *Difusão Lexical: estudo de casos do Português Brasileiro*. Relatório de Pesquisa CNPq número: 202422-86-3. Manuscrito. Belo Horizonte: Faculdade de Letras da UFMG, 2003.

FONTES MARTINS, Raquel M. *O cancelamento das líquidas “l” e “r” intervocálicas no português de Belo Horizonte*. 2001. Dissertação (Mestrado em Linguística) – Faculdade de Letras, UFMG, Belo Horizonte, 2001.

GUIMARÃES, Daniela Mara Lima. *Variação nas seqüências de (sibilante +africada alveolopalatal) no português de Belo Horizonte*. 2004. Dissertação (Mestrado em Linguística) – Faculdade de Letras, UFMG, Belo Horizonte, 2004.

HAZEN, Kirk. The Family. In: CHAMBERS, J. K.; TRUDGILL, Peter; SCHILLING-ESTES, Natalie (Ed.). *The Handbook of Language Variation and Change*. Oxford: Blackwell, 2002. p. 500-525.

HUBACK, Ana Paula da Silva. *Cancelamento do (R) final em nominais na cidade de Belo Horizonte: uma abordagem difusionista*. 2003. Dissertação (Mestrado em Linguística) – Faculdade de Letras, UFMG, Belo Horizonte, 2003.

JOHNSON, K.; MULLENNIX (Ed.). *Talker variability in Speech Processing*. San Diego: Academic Press, 1997.

LABOV, W. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press, 1972a.

LABOV, W. Exact description of the speech community: Short *a* in Philadelphia. In: FASOLD, Ralph W.; SCHIFFRIN, Deborah (Eds.). *Language Change and Variation*. Washington, DC: Georgetown University Press, 1989b. p. 1-57.

LABOV, W. *Principles of linguistic change: social factors*. Oxford: Blackwell, v. 2, 2001.

LINGÜÍSTICA APLICADA E ESTUDOS DA LINGUAGEM (LAEL). *Corpus oral e escrito da língua portuguesa*. Disponível em: <<http://lael.pucsp.br/corpora/index.htm>>. Acesso em: fev. e mar. 2004.

MARSHALL, Jonathan. *Language change and sociolinguistics: rethinking social networks*. New York: Palgrave Macmillan, 2004.

NARO, Anthony. O dinamismo das línguas. In: MOLLICA; BRAGA (Org.). *Introdução à sociolinguística: o tratamento da variação*. São Paulo: Contexto, 2003, p. 43-50.

OLIVEIRA, Marco Antônio de. Aspectos da difusão lexical. *Revista de Estudos da Linguagem*. Belo Horizonte: Faculdade de Letras da UFMG, v. 1, p. 31-41, 1992.

OSTHOFF; BRUGMAN. *Preface to morphological investigations in the sphere of the Indo-European languages*, [s.l.], 1878. p. 299-209.

PAIVA, Maria da Conceição; DUARTE, Maria Eugênia Lamoglia (Org.). *Mudança linguística em tempo real*. Rio de Janeiro: Contra-Capa, 2003.

PAGOTTO, Emílio Gozze. *Variação e (') identidade*. Maceió: UFAL, 2004.

PIERREHUMBERT, Janet. Exemplar dynamics: Word frequency, lenition, and contrast. In: BYBEE, J.; HOPPER, P. (Ed.). *Frequency effects and the emergence of linguistic structure*. Amsterdam: John Benjamins, 2001. p. 137-157.

REINECKE, Katja. *Os róticos intervocálicos na gramática individual de falantes de Blumenau e Lages*. A sair. Tese (Doutorado em Linguística) – Faculdade de Letras, UFSC, Florianópolis. A sair.

ROCHA, Luiz Carlos de Assis. *Estruturas morfológicas do português*. Belo Horizonte: UFMG, 1998.

SANGSTER, Catherine M. *Inter- and Intra-Speaker Variation in Liverpool English: A Sociophonetic Study*. 2002. 290 f. Tese de Doutorado – University of Oxford, Oxford, 2002.

WEINREICH, Uriel; LABOV, William; HERZOG, Marvin. Empirical foundations for a theory of language change. In: LEHMANN, W. P.; MALKIEL, Yakov (Eds.). *Directions in Historical Linguistics*. Austin, TX:University of Texas Press, 1968. p. 95-188.

## ANEXOS

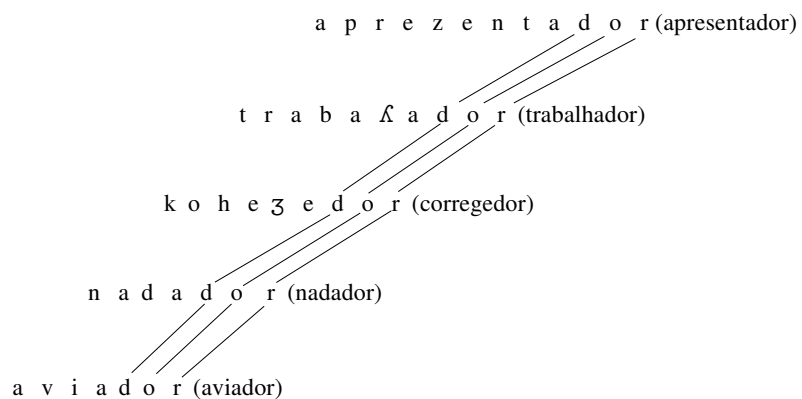


FIGURA 1 – Esquema com conexões lexicais (fonológicas e semânticas – relações morfológicas) para o morfema *-dor*

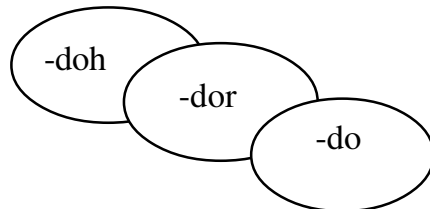


FIGURA 2 – Nuvem de exemplares para o sufixo *-dor*

*\*Baseado em Cristófar-Silva (2002)*

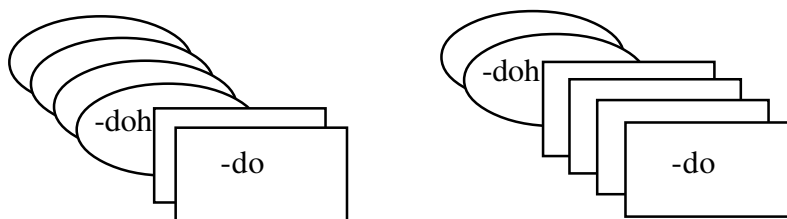


FIGURA 3 – Nuvem de exemplares para o sufixo *-dor*, com *-doh* e *-do* em competição

*\*Baseado em Cristófar-Silva (2002)*

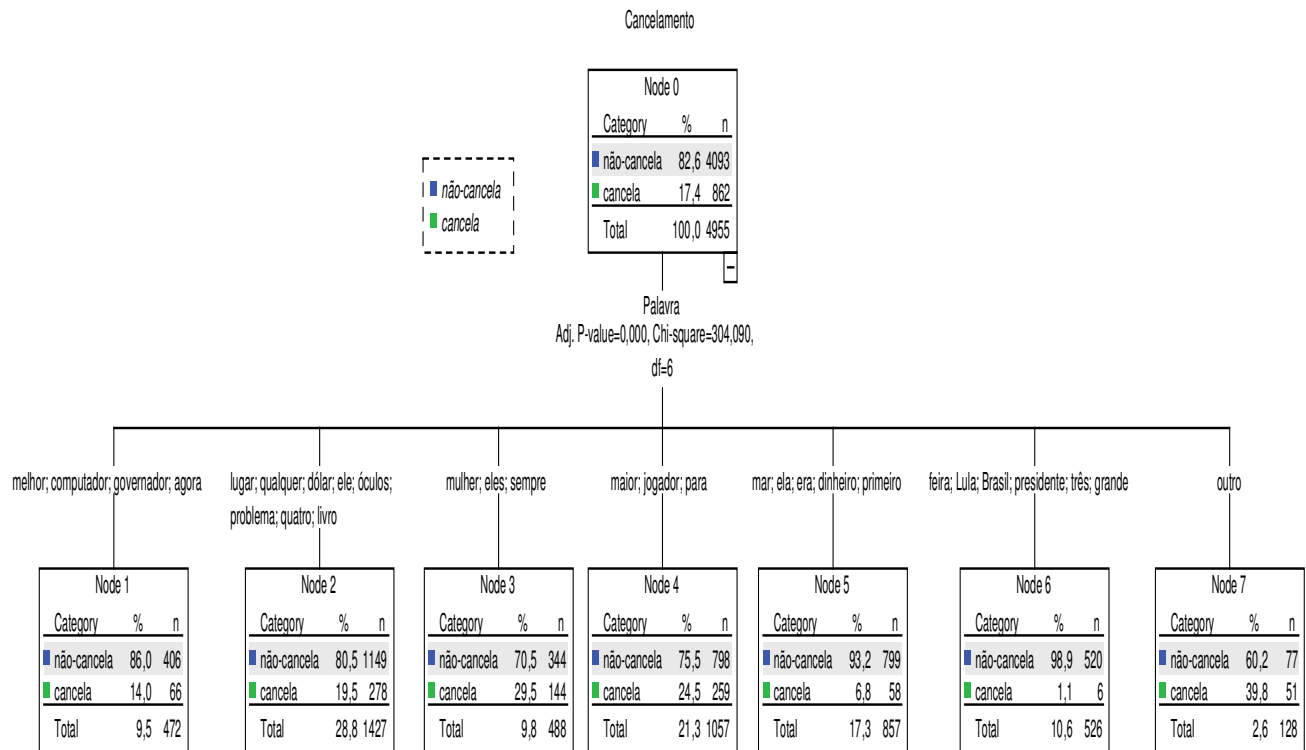


FIGURA 4 – Árvore (método Chaid) com classificação para os dados do estudo transversal (4.955 ocorrências para 30 tipos de palavras)

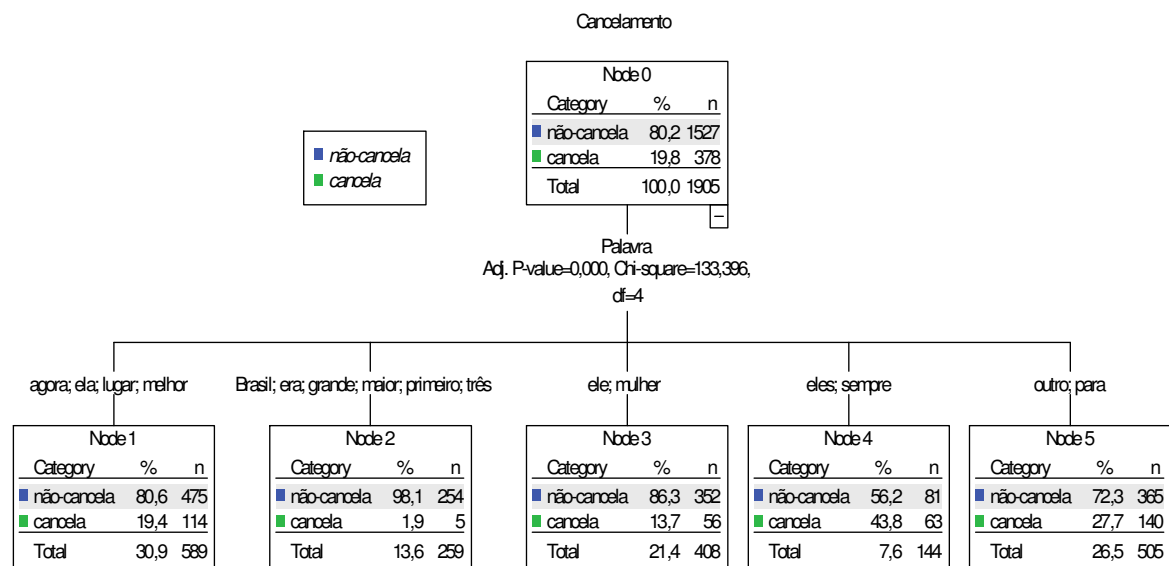


FIGURA 5 – Árvore (método Chaid) com classificação para os dados do estudo longitudinal (1.905 ocorrências para 16 tipos de palavras)

QUADRO 1  
30 palavras – estudo transversal

<i>Corpora</i>	Palavras	Fenômeno
<i>CORPUS 1:</i> Huback (2003)	1. Melhor	Cancelamento de “r” em final de palavra
	2. Lugar	
	3. Mulher	
	4. Qualquer	
	5. Computador	
<i>CORPUS 2:</i> Fontes Martins (2001)	6. Para	Cancelamento de líquida intervocálica
	7. Ele	
	8. Ela	
	9. Eles	
<i>CORPUS 3:</i> Cristófaró-Silva (2003)	10. Óculos	Cancelamento de encontro consonantal tautossilábico
	11. Problema	
	12. Outro	
	13. Quatro	
	14. Sempre	
LAEL -escrita	15. Primeiro	Cancelamento de “r” em final de palavra
	16. Maior	
	17. Mar	
	18. Governador	
	19. Jogador	Cancelamento de líquida intervocálica
	20. Dólar	
	21. Era	
	22. Dinheiro	
	23. Feira	Cancelamento de encontro consonantal tautossilábico
	24. Agora	
	25. Lula	
	26. Brasil	
	27. Presidente	
	28. Três	
29. Grande		
30. Livro		

